THE SPECTRAL METHOD: ITS IMPACT ON NWP

Ferdinand Baer
Department of Meteorology
University of Maryland
College Park, MD 20742 USA

Weather forecasting was first presented as a computational problem by Richardson (1922) based on the equations of fluid dynamics, and started its meteoric rise with the advent of digital computers. The approximations needed to convert the relevant differential equations to numerical form suitable for computation on large computing machines thus evolved. From this perspective, the concept of modeling was conceived; an atmospheric numerical 'model' was developed and solutions to this model were sought. A variety of models have been developed since to meet this goal, and the 'spectral model' is one of these.

For this purpose the atmosphere is represented by variables describing molecular composites of its gases; the primary variables are velocity, temperature, density, water content in all phases, and aerosols. These variables are considered to be distributed continuously in three-dimensional space surrounding the Earth and vary with time. The time evolution of these variables may be determined at each point in space (the Eulerian method) or by following the particles through time (the Lagrangian method) and both methods are in use. The differential equations defining the future state of the variables are based on physical and dynamical principles. These principles include the equations of motion (the Navier-Stokes equations), conservation of mass, an equation for change in entropy, equations for changes in water substance in its various phases, and chemical equations for changes of aerosols.

In general terms, suppose a vector $\mathbf{B}$ to represents N dependent variables describing the state of the prediction system at any time such that $\mathbf{B} = \{B_b\} = (\mathbf{V}\ \rho\ s\ q_v\ q_i\ q_l\ a_j \cdots)^T$ where T stands for transpose. The weather prediction system can be written as

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{F}(\mathbf{B}, \mathbf{r}, t) \tag{1}$$

where $\mathbf{F}$ has the dimensions of $\mathbf{B}$, depends nonlinearly and differentially on $\mathbf{B}$, and depends on $\mathbf{r}$ (the space coordinates) and t. A linear matrix operator $\mathbf{L}$ is often used to transform these equations,

$$\mathbf{L}\frac{\partial \mathbf{B}}{\partial t} = \tilde{\mathbf{F}}(\mathbf{B}, \mathbf{r}, t) \tag{2}$$

This system of equations is the basis from which the 'model' is selected and integrated numerically in time to predict the future state of the fluid. Additional features needed to complete the model are boundary conditions, initial conditions, scale truncation, external forces and computational resources.

The 'model' is constructed by selecting a methodology to convert the basic nonlinear system (2) to a numerical

form suitable for computation and integration on a digital computer. Finite differencing in both the time and space dimensions was the first method attempted and is still successfully used for many applications. Since the vertical and horizontal dimensions in the atmosphere have unique properties, they often are considered separately. Given that at any given height in the atmosphere a closed spherical surface exists on which the dependent variables describing the fluid are prescribed and predicted, the spectral method, which assigns a set of known continuous orthogonal functions over the domain to represent these variables has also been utilized. When all the variables are described in this way, the resulting system is integrated over the global domain, leading to a set of ordinary nonlinear differential equations in time and on each vertical level. Concurrently, differentiation in the vertical space coordinate and time is generally transformed to finite differences. The spectral method is most appropriate for the larger space scales since the functions usually used are global. Additional methods that have been utilized include finite-elements and spherical geodesic grids. Let us compare the finite difference system to the spectral system to evaluate their relative properties.

1. THE FINITE DIFFERENCE METHOD

To approximate the continuum in space, choose a three-dimensional grid with M points and suitably prescribed boundary conditions, and a difference operator to evaluate derivatives. Assigning values to **B** at each of the points and using them as a set of initial values, a numerical integration can proceed. Note that **B** has dimensions (NxM). The matrix **L** becomes, by virtue of the difference operator, an (NxM) x (NxM) matrix that can in principle be inverted, and **F** as well becomes a numerical vector with NxM elements after application of the difference operator at each grid point. The final finite difference system is then written, using a circumflex to represent the numerical vectors and matrices at the grid points,

$$\frac{\partial \hat{\mathbf{B}}}{\partial t} = \hat{\mathbf{L}}^{-1}\hat{\mathbf{F}}\left(\hat{\mathbf{B}},\hat{\mathbf{r}},t\right) \tag{3}$$

The solution is thus reduced to a matrix computation provided a numerical scheme is introduced to step the solution forward in time, and the resulting computational errors and stability issues are explicit in the numerical and physical approximations made.

2. THE SPECTRAL METHOD

On the continuous domain over which the model variables are to be evaluated, select a set of linearly independent global functions (Z_m) that are continuous over the domain and have at least continuous first and second derivatives. The model variables B_b are expanded in these functions with unknown time dependent coefficients. Thus instead of a set of values for the B_b at each grid point ($i\Delta x_1$, $j\Delta x_2$, $k\Delta x_3$) one has

$$B_b = \sum_{m=1}^{M_e} B_{b,m}(t) Z_m(\mathbf{r}) \tag{4}$$

For application to the prediction system introduce (4) into (2). To maintain the exact form of (2) the series given by (4) must be infinite. Using a truncated form will cause an error, just as the reduction to a grid does in (3). Selecting an optimum truncation is a

significant issue. The choice of these functions is arbitrary but some guidelines may optimize their selection. It would be ideal to select functions that fit the observation points of the expanded variables exactly but the distribution of observations is not sufficiently uniform to make this feasible. The expansion functions might be chosen to fit statistics of observations interpolated to a more uniform grid such that the least number of functions (N) were required to describe most of the variance of the variables at those points. Additionally, functions could be chosen that fit boundary conditions most efficiently and/or with convenient orthogonalization properties.

The operator **L** is often used with the spectral method. This system is always linearly decoupled, so **L** is represented as a diagonal matrix with L_b elements on the diagonal. The prediction equation for each variable b in scalar form then appears as,

$$L_b \frac{\partial B_b}{\partial t} = \tilde{F}_b(\mathbf{B}, \mathbf{r}, t) \tag{5}$$

Note the variables are still nonlinearly coupled in the functions. Substitution of (4) into (5) leads to the error equation,

$$\sum_{a=1}^{M_e} \left(\frac{\partial B_{b,a}}{\partial t} L_b Z_a \right) - \tilde{F}_b = \varepsilon_b \tag{6}$$

To solve this system for the unknown expansion coefficients $B_{b,a}$, multiply (6) by suitable test functions $\hat{Z}_k(\mathbf{r})$ and require the integral over the space domain to vanish. These test functions must be continuous over the domain, and can be arbitrary, but are often chosen to be the expansion functions. The prediction equations for the expansion coefficients then become,

$$\sum_{a=1}^{M_e} \left(\frac{\partial B_{b,a}}{\partial t} \int L_b Z_a \hat{Z}_k dS \right) - \int \tilde{F}_b \hat{Z}_k dS = 0 \tag{7}$$

and yield NxM_e equations for the unknown quantities, $\partial B_{b,a} / \partial t$. This system can be solved by choice of a suitable time extrapolation procedure.

To represent (7) in a form more comparable to the finite difference equations (3), let $\mathbf{B}_b = (B_{b,a})$ and $\mathbf{Z} = (Z_a)$, both vectors with M_e elements and the test functions as $\hat{\mathbf{Z}} = (\hat{Z}_k)$. Recalling that the functions F_b are implicitly functions of (**r**, t), projection onto the expansion functions yields,

$$F_b = \sum_a F_{b,a}(t) Z_a = \mathbf{Z}^T \mathbf{F}_b \tag{8}$$

where $\mathbf{F}_b = (F_{b,a})$. Using the defined vectors, (7) becomes,

$$\int \hat{\mathbf{Z}} L_b \mathbf{Z}^T dS \cdot \frac{\partial \mathbf{B}_b}{\partial t} = \int \hat{\mathbf{Z}} \mathbf{Z}^T dS \cdot \mathbf{F}_b \tag{9}$$

and represents M_e equations for the expansion coefficients of each dependent variable. To combine the N equations of (9) into one expression, it is convenient to define the M_e x M_e matrices $\mathbf{A}_b \equiv \int \hat{\mathbf{Z}} L_b \mathbf{Z}^T dS$ and $\mathbf{A} \equiv \int \hat{\mathbf{Z}} \mathbf{Z}^T dS$, and then create $(NxM_e)x(NxM_e)$ matrices having these matrices on the diagonals; i.e., $\mathbf{A}_L$ = diag ($\mathbf{A}_b$) and $\mathbf{A}_R$ = diag ($\mathbf{A}$). Extended vectors for the expansion coefficients to include all the variables can be constructed such that $\mathbf{B}_s$ = ($\mathbf{B}_b$) and $\mathbf{F}_s$ = ($\mathbf{F}_b$), leading finally to an equation which is formally identical to the finite difference equation (3),

$$\frac{\partial \mathbf{B}_s}{\partial t} = \mathbf{A}_L^{-1} \mathbf{A}_R \mathbf{F}_s \tag{10}$$

The corresponding grid point values from this spectral representation are calculated at each point ($i\Delta x_1$, $j\Delta x_2$, $k\Delta x_3$) for each dependent variable B_b by

use of (4) and can then be compared to the values derived from the Finite Difference equation (3). This comparison will yield information on how each method deals with its own truncation and when comparisons are made to observations, one can assess the relative qualities of each method.

2.1 Specifics of the Spectral method used for NWP

The subsequent discussion will focus on the horizontal domain of the model representation because most significant prediction models represent their dependent variables on a grid of points in the vertical and use non-spectral methods on that grid. This requires that the variables B_b be represented on K surfaces in the vertical, with the surfaces separated by the grid intervals, and the variables described in those surfaces by (4).

When selecting appropriate spectral functions for the expansion (4), in addition to fitting observations well, the functions should also be chosen with the properties of the system in mind. Several of these properties follow. First require the functions Z_a to satisfy the eigenvalue problem,

$$L_b Z_a = -c_{b,a} Z_a. \qquad (11).$$

In practice the selection of L_b almost always represents a conversion of wind components to vorticity and divergence, which is given by a linear differential operator. Application of this operator in (11) leads to a variety of useful and simple functions. The second condition is to require the expansion functions to be orthogonal and normal over the domain in a Hermitian sense,

$$\int Z_i Z_j^* dS = \delta_{i,j} \qquad (12).$$

This condition is reasonably simple to satisfy, since most function sets can be orthogonalized. Finally, the test functions when selected as the expansion functions do not lead to a significant loss of generality, thus this condition is uniformly imposed as $\hat{\mathbf{Z}} = \mathbf{Z}$. Utilizing these three conditions greatly simplifies the calculations required to perform each prediction time step since both integrals in (9) become diagonal matrices.

A variety of functions have been used for the expansion (4), most satisfying the conditions just enumerated, with the selection depending on the degree of generality desired to approximate the general system (10). When the atmosphere is represented on a channel with rigid boundaries at fixed northern and southern latitudes short of the poles, double Fourier series in latitude and longitude are found to be convenient expansion functions. They satisfy the boundary conditions easily, and because of the very simple addition rules for these functions, nonlinear products are rapidly calculated. For the full global domain approximated by spherical surfaces over the Earth, the obvious expansion functions that satisfy the boundary conditions are surface spherical harmonics (often denoted as solid harmonics), and they have become the functions of choice for spectral modeling. Surface spherical harmonics are constructed as the product of associated Legendre polynomials and complex exponential functions. Selecting coordinates in spherical surfaces such that $\mu = \sin\varphi$, where φ is latitude, and λ is longitude, normalized Legendre polynomials represent the latitudinal structures with the following form;

$$P_n^m(\mu) \equiv \left[(2n+1)\frac{(n-m)!}{(n+m)!}\right]^{1/2} \times \frac{(1-\mu^2)^{m/2}}{2^n n!}\left(\frac{d}{d\mu}\right)^{n+m}(\mu^2-1)^n \quad (13).$$

These are polynomials of degree n with n-m roots in the domain $-\pi/2 < \varphi < \pi/2$ and m roots at the poles. Together with Fourier series in longitude the solid harmonics are,

$$Y_{n,m}(\lambda, \mu) = P_{n,m}(\mu)e^{im\lambda} \quad (14).$$

These are the complex expansion functions Z_a used in (4) for the horizontal structures. All functions vanish at the poles except the zonal ones (m = 0), and these remain finite there. The indices (n, m) define the roots of the functions and thus may be considered scaling elements; i.e., the larger the indices, the smaller the scales represented by the functions. An example is given in Figure 1, which shows the cellular structure of the function for fixed n and various values of m. The total number of cells over the domain remains the same since some of the roots appear at the poles, but the cell structures differ. It is convenient to represent the indices as a single complex index, say $\alpha = (n + im)$. The functions are orthogonal over their respective domains and normalized, and this is expressed (in a Hermitian sense) as

$$\frac{1}{4\pi}\int Y_\alpha Y_{\alpha'}^* dS = \delta_{\alpha,\alpha'} \quad (15),$$

with integration taken over the surface of the unit sphere. The asterisk signifies complex conjugation, and δ is the Kroniker delta. If $L_b \equiv \nabla^2$ (the Laplacian operator), substitution of Y_α for Z_a in (11) yields eigenvalues,

$$c_\alpha = n(n+1) \quad (16).$$

Thus solid harmonics satisfy the conditions desired for suitable expansion functions.

Most atmospheric variables (B_b) are sufficiently smooth that when expanded in these functions, the series converges rapidly. That expansion takes the form,

$$B_b(\lambda,\mu,z_k,t) = \sum_\alpha B_{b,\alpha,k}(t)Y_\alpha(\lambda,\mu) \quad (17),$$

where z_k is any selected vertical level and the series truncates at M_e. The range of (α) is $n \geq 0$ and because of the complex nature of Fourier series, m takes both positive and negative values. When (17) is introduced into (10) and suitably integrated over the space domain, the resulting equations become a set of ordinary nonlinear differential equations in time for the expansion coefficients.

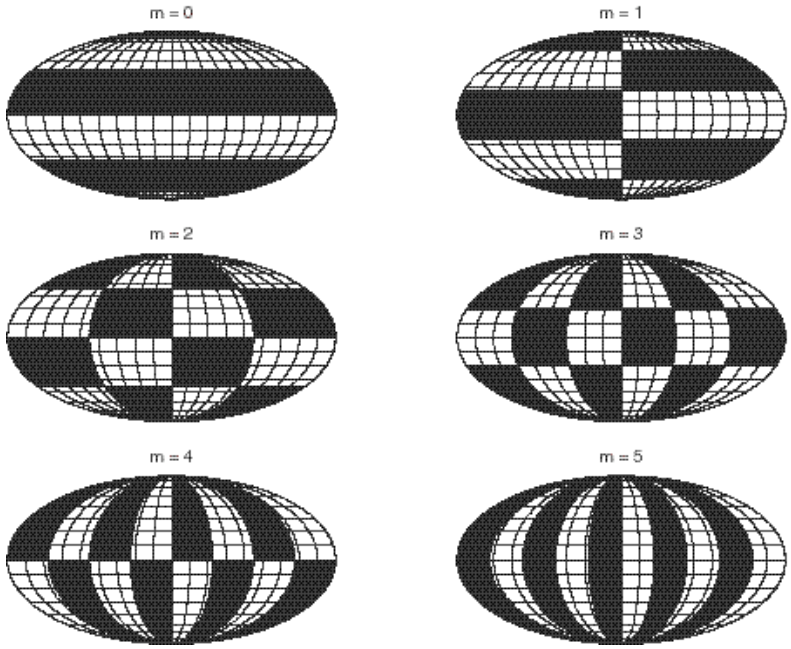

Figure 1. Cellular structure of solid harmonic functions for n =5 and all allowed values of m.

2.2 The barotropic vorticity equation

To understand some of the advantageous features of the spectral method it is convenient to use a simple reduction of the complete system for demonstration purposes. Consider a barotropic fluid, which exists if the thermodynamic variables are uniquely related to one another and are independent of position in the fluid. Since the vertical variation becomes irrelevant to the solution of the fluid motions in this setting, the motions need consideration in only one horizontal surface. In this context it is convenient to assume that the fluid is incompressible, which leads to three-dimensional non-divergence. If no divergence is introduced at the upper and lower boundaries, the continuity equation can be integrated over the vertical domain to demonstrate that no divergence exists in any horizontal surface. Finally, if the fluid is considered to be in hydrostatic equilibrium, it is only necessary to establish the evolution of the horizontal velocity in any such surface and the vertical component of velocity can be ignored. This velocity is represented by two scalar variables, which may be transformed to any other two scalar functions. Because rotation plays such a major role in atmospheric motions, vorticity and divergence are the popular choices for this transformation, and since the divergence vanishes for this discussion the velocity may be represented uniquely by the vorticity alone. This leads to the Barotropic Vorticity Equation (BVE). This equation, although representing a very simplified version of the atmosphere, contains many features of the full atmospheric system and is thus a convenient tool for evaluating prediction techniques.

Applying the approximations stated above and ignoring friction, the resulting equation of motion may be written;

$$\frac{\partial \mathbf{V_2}}{\partial t} = -\mathbf{V_2}\cdot\nabla\mathbf{V_2} - 2\vec{\Omega}\times\mathbf{V_2} - \frac{1}{\rho(p)}\nabla_2 p \quad (18),$$

where the subscript 2 denotes two-dimensionality. The Earth's vorticity may be expressed here as $2\vec{\Omega} = f\mathbf{k}$ where $f = 2\Omega \sin\varphi$, the Coriolis parameter, and φ is latitude. If the velocity is now transformed to rotation and divergence by the definitions,

$$\mathbf{V_2} = \mathbf{k}\times\nabla\psi + \nabla\chi$$
$$\nabla\cdot\mathbf{V_2} = \nabla^2\chi \equiv \delta \equiv \textit{divergence}$$
$$\mathbf{k}\cdot\nabla\times\mathbf{V_2} = \nabla^2\psi \equiv \zeta \equiv \textit{relative vorticity} \quad (19),$$

an equation for predicting the vorticity may be established by applying the operator $\mathbf{k}\cdot\nabla\times$ to (18);

$$\mathbf{k}\cdot\nabla\times\left[\frac{\partial \mathbf{V_2}}{\partial t} = -\mathbf{V_2}\cdot\nabla\mathbf{V_2} - f\mathbf{k}\times\mathbf{V_2}\right] \quad (20).$$

Noting that the divergence vanishes, and substituting the definitions from (19), the barotropic vorticity equation emerges as,

$$\frac{\partial\zeta}{\partial t} = -V_2\cdot\nabla\eta = -\mathbf{k}\times\nabla\psi\cdot\nabla\eta = -J(\psi,\eta)$$
$$\eta \equiv \zeta + f \equiv \textit{absolute vorticity} \quad (21).$$

If one nondimensionalizes (21) using the Earth's radius (a) for space and its rotation rate (Ω) for time and noting that the Coriolis parameter is then $f = 2\mu$, the vorticity may be written in terms of the stream function (ψ) as,

$$\frac{\partial\nabla^2\psi}{\partial t} = -2\frac{\partial\psi}{\partial\lambda} - F(\psi)$$
$$F(\psi) \equiv \frac{\partial\psi}{\partial\lambda}\frac{\partial\nabla^2\psi}{\partial\mu} - \frac{\partial\psi}{\partial\mu}\frac{\partial\nabla^2\psi}{\partial\lambda} \quad (22).$$

Indeed, $\psi = B$, the only variable remaining of the set B_b and for only one k level. Eq. (22) contains a linear term and two quadratic nonlinear terms; these terms constitute F, what remains of F_b in (8). The representation in terms of expansion coefficients ψ_α (t) can be attained using (11) and $L_b \equiv \nabla^2$, (17) for the expansion of ψ, and a similar for expansion of F, to yield,

$$\sum_\alpha c_\alpha Y_\alpha(\lambda,\mu)\frac{\partial\psi_\alpha(t)}{\partial t} = 2i\sum_\alpha m_\alpha\psi_\alpha Y_\alpha(\lambda,\mu) + \sum_\alpha F_\alpha Y_\alpha(\lambda,\mu) \quad (23).$$

The final step is to multiply (23) by the test functions and integrate over the unit sphere. In this case the solid harmonics are themselves the test functions and to exploit orthogonality the product is with the conjugate of each harmonic. This results in the prediction equation for each of the expansion coefficients,

$$\frac{\partial\psi_\alpha(t)}{\partial t} = 2im_\alpha c_\alpha^{-1}\psi_\alpha(t) + F_\alpha(t)$$
$$F_\alpha(t) = \int F(\psi)Y_\alpha^*(\lambda,\mu)dS \quad (24).$$

It should be apparent how (24) can be extended to involve more dependent variables and any number of levels in the vertical. However, if more variables exist in the system, these variables will be coupled nonlinearly through the coefficients F_α.

2.3 Computational features of the spectral method

Interaction coefficient method

Since all prediction models are computationally intensive, the spectral method must compete in the efficient utilization of available computing resources. We shall use the BVE as a demonstration model in the following. It is apparent from (24) that most of the computing time required involves the calculation of the coefficients F_α and much effort has gone into optimizing this calculation. Early attempts followed the procedure of substituting the

expansion series (17) for ψ into (22) to represent $F(\psi)$ and calculating F_α from (24). This results in,

$$F_\alpha(t) = \frac{i}{2}\sum_\beta \sum_\gamma \psi_\beta(t)\psi_\gamma(t) I_{\alpha,\beta,\gamma}$$

$$I_{\alpha,\beta,\gamma} \equiv (c_\beta - c_\gamma) \times \qquad (25).$$

$$\int \left(m_\beta Y_\beta \frac{\partial Y_\gamma}{\partial \mu} - m_\gamma Y_\gamma \frac{\partial Y_\beta}{\partial \mu} \right) Y_\alpha^* dS$$

The indices β and γ go over the same range as α which is determined by the selected truncation and the integration is over the unit sphere. The integrals $I_{\alpha,\beta,\gamma}$ are denoted as *interaction coefficients* and have exact solutions. Applying (25) in (24) shows that the time change of any expansion coefficient of the set α depends on the coupling of all the coefficients allowed in the spectral domain (refer to Figure 2) and each couple is weighted by its own interaction coefficient. Since each index consists of two real numbers, the set of interaction coefficients can be as large as the largest allowed index to the sixth power. In actuality, because of the simple addition rules for trigonometric functions, the integration over longitude reduces this by one order. The vector of these coefficients can be stored and need be computed only once. However the number of multiplications that must be performed at each time step is daunting as the truncation limit becomes large.

The more complex system (10) can be represented identically to (24) by simply increasing the number of expansion coefficients to include additional variables. But a shortcoming of using interaction coefficients concerns the convergence rate for the series of several dependent variables included in the general set (B_b) when expanded in global functions, in particular liquid water and precipitation. Significant truncation errors may ensue with time integration utilizing such functions.

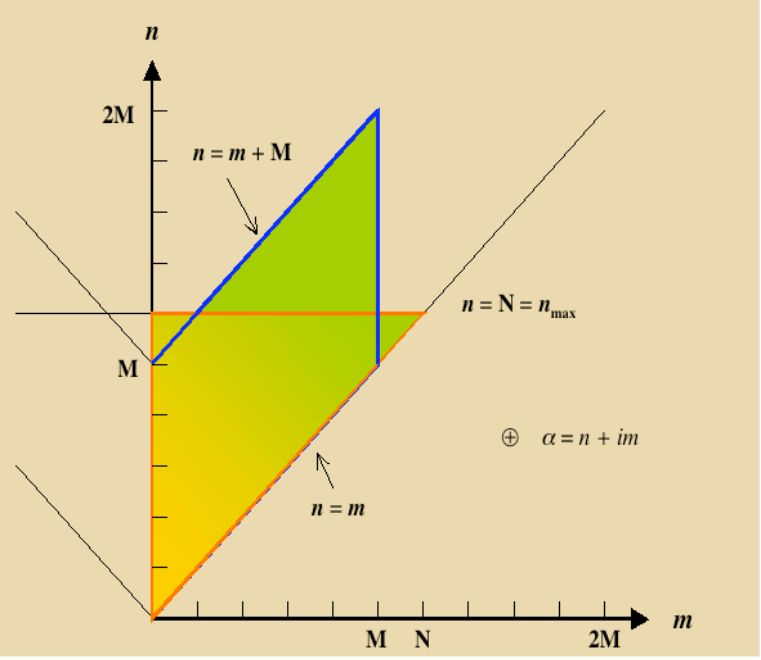

Figure 2: Truncation diagram in a spherical surface using the spectral method. Several popular truncations are depicted.

Transform method

A technique denoted as the *transform method* is an alternate procedure for calculating the coefficients $F_{\alpha,}$ yielding the same (or better) results than the interaction coefficient method. This technique involves the transformation of the integrand in (24) onto a special numerical grid and solving the integral by quadrature. If the grid is selected appropriately, the integral is evaluated exactly and at a great reduction in computing cost. In the longitudinal direction the quadrature is most conveniently done by a trapezoidal formula since it is known that

$$\frac{1}{2\pi}\int_0^{2\pi} e^{im\lambda}\,d\lambda = \frac{1}{J}\sum_{j=1}^{J} e^{im\lambda_j} \qquad (26).$$

The summation is taken over an equally spaced grid of points λ_j, and uses twice the number of points as the maximum wave number. Since the functions in latitude are Legendre polynomials, a Gaussian quadrature is preferred. In this case the quadrature is such that,

$$\int_{-1}^{1} H(\mu)d\mu = \sum_{k=1}^{K} G_k(\mu,K)H(\mu_k) \quad (27)$$

and is exact if the polynomial H is of degree 2K-1 or less. The G_k are Gaussian weights and the grid points μ_k are the roots of the Legendre polynomial $P_K(\mu)$. The appropriate grid for this calculation contains all allowed values ($\lambda_{j,}$ μ_k) as specified. The range of the grid points is determined by the functions of the integrand in (24).

The derivatives in F(ψ), see (22), must be taken before evaluating the function on the grid. Based on (14) and (13) the differentiation with λ is straightforward, but the μ-derivative requires more information. The Legendre polynomials satisfy the following differential equation,

$$(1-\mu^2)^{1/2}\frac{dP_\alpha}{d\mu} = b_\alpha P_{\alpha-1} - b_{\alpha+1}P_{\alpha+1} \qquad (28)$$

where the coefficients b_α are constants, and this defines the latitudinal derivatives. Following this procedure, F(ψ) is reduced to a quadratic series over the indices (β, γ) in terms of the complex exponential functions in longitude and the associated Legendre polynomials in latitude, both of which can be evaluated on the specified grid. The actual calculation proceeds as follows. First the quadrature over longitude is taken,

$$F_{m_\alpha}(\mu,t) = \frac{1}{2\pi}\int F\left(\psi(\lambda,\mu,t)\right)e^{-im_\alpha\lambda}d\lambda = \frac{1}{J}\sum_{j=1}^{J} F\left(\psi(\lambda_j,\mu,t)\right)e^{-im_\alpha\lambda_j} \qquad (29)$$

where the sum goes over the value J = 3M -1 if triangular truncation is chosen. The calculation is made over those latitudes μ specified from the quadrature,

$$F_\alpha(t) = \frac{1}{2}\int F_{m_\alpha}(\mu,t)P_\alpha(\mu)d\mu = \sum_{1}^{K} G_k(\mu_k,K)F_{m_\alpha}(\mu_k,t)P_\alpha(\mu_k) \qquad (30).$$

Since the polynomial under summation in (30) is H(μ) and is the product of three Legendre polynomials less one order, and each has a maximum order of N, it can be shown that K = (3N-1)/2.

Analysis of the computing requirements for (29) and (30) indicates that the maximum number of calculations is proportional to N^3, significantly less than the N^5 needed by

the interaction coefficient method. When using the transform method with those variables that have unacceptable convergence properties yet contribute to (3), their series representation is not essential. Their input is included directly into the quadrature formula by their distribution on the transform grid. Since all the forcing functions are summed over the grid before quadrature is completed, any singularities from individual terms are smoothed out and their effects are minimized. It was the benefits of the transform method that made the spectral method a popular alternative to the finite difference method.

2.4 Advantageous features of the spectral method

Linear instability and phase errors

It can be seen from (24) that the prediction equation is made up of a linear and a nonlinear term. Since this is the BVE, only one dependent variable exists and it is clear that the linear term can be dealt with exactly by solving only the linear part of the equation. The dependent variable can be transformed by that solution and the resulting equation will contain only a nonlinear term. Thus linear instability and associated phase propagation errors can be completely avoided. For the more general system (10) the dependent variables constitute a vector and thus the linear system would look like,

$$\frac{\partial \mathbf{B}_s}{\partial t} = \mathbf{C}\mathbf{B}_s + \text{nonlinear terms} \qquad (31),$$

where **C** is a matrix of constant coefficients. This matrix may be diagonalized and the linear part of the equation may be solved. Transforming $\mathbf{B}_s$ appropriately, the linear terms can be removed from the equation and only the nonlinear terms need be calculated, again removing the possibility of potential linear instability and phase errors. However, many studies have shown that this is not true for finite difference equations. The errors that result from linear wave dispersion can have a dramatic effect when they are included in the nonlinear wave interactions, setting up systematic errors as the time integration evolves.

Nonlinear instability

Phillips (1959) brought the problem of nonlinear instability to the modeling community's attention, an issue requiring careful attention lest the evolving solutions become unstable. Finite difference equation (3) shows that nonlinear terms of the form $u\,\partial u/\partial x$ exist and must be calculated. Assume for simplicity that u is one of the velocity components and is a function of only one space dimension (x). Since the variables must be represented on a grid of points truncated at some Δx, one could as well represent u as a Fourier series with the shortest resolvable wave length of $2\Delta x$ and corresponding wave numbers $m \geq \pi/\Delta x$. If only sine functions are considered, u can be written as,

$$u = \sum_{i=1}^{M} u_i \sin m_i x \qquad (32)$$

over M grid points. The nonlinear product of any two waves, m_I and m_j becomes,

$$u\,\partial u/\partial x = m_j \sin m_i x \cos m_j x + \cdots = \tfrac{1}{2} m_j \left[\sin\left(m_i + m_j\right)x + \sin\left(m_i - m_j\right)x\right] + \cdots \qquad (33).$$

Consider now that u can be predicted from this term and others as,

$$u(t + \Delta t) = u(t) + \Delta t\,(u\,\partial u/\partial t) + \cdots \quad (34).$$

So long as $m_I + m_j \leq M$, $u(t + \Delta t)$ can be represented by the Fourier series in (32) without error. However if the sum of the two wave numbers exceeds M, the series limit is exceeded. An expansion of $u(t + \Delta t)$ with M terms will not see this interaction. Worse yet, it can be shown by trigonometric expansion that terms larger than $M\Delta x$ will fold back into the range $(M-m_k)\,\Delta x$ creating an aliasing error that can grow without bound and is thus denoted as nonlinear instability. Since this error tends to affect the shortest scales, the solution has been to include in the model equations a scale sensitive viscosity to damp these scales before they grow significantly, a rather arbitrary and nonphysical, but practical, decision.

Interestingly, the spectral form of the equations is insensitive to this problem. Suppose that the series for α in (24) is truncated at M_e as suggested. This implies that all values of ψ_α for $\alpha > M_e$ vanish. However on calculating the nonlinear product $F(\psi)$ one could get coefficients F_α for $\alpha \leq 2M_e$ since the quadratic product of a polynomial will yield a polynomial with twice the maximum order. Thus in principle at each time step, the number of non-vanishing coefficients could double. This complication is uniformly resolved in the spectral method by ignoring all computations for $\alpha > M_e$ once a calculation has begun. Since the products with $\alpha > M_e$ are discarded, they cannot fold back into the domain of $\alpha < M_e$ to corrupt the coefficients in that range. This allows for stable computations with no requirement for artificial viscosity, regardless of the complexity of the prediction system.

Conservation

A prominent feature of the atmospheric prediction system is that in the absence of external energy sources, the total energy of the system must be conserved in time. Despite the sources that come from say radiation and surface boundary effects, it is important that the conservation conditions inherent in the differential equations be maintained in the computational equations. Setting up the finite difference equations to conserve energy is a daunting task and has engaged the best researchers for decades. It is particularly difficult because of the three dimensional nature of the task. The significance of conservation in a truncated framework is most simply seen by use of the BVE. It is an easy matter to demonstrate that based on (21) the absolute vorticity or any function thereof must be conserved both following a particle and on integration over the surface of all particles. When this equation is converted to finite difference form, the condition of conservation should also be met.

Arbitrary application of a finite difference operator to (21) will not achieve this required result. Arakawa (1966) presented a procedure that assures conservation not only of vorticity, but kinetic energy and enstrophy (mean squared vorticity) as well. The technique involved multiple representations of the Jacobian operator; Arakawa used the following three, with x and y as the independent variables and η as absolute vorticity;

$$J(\psi,\eta) = \frac{\partial\psi}{\partial x}\frac{\partial\eta}{\partial y} - \frac{\partial\psi}{\partial y}\frac{\partial\eta}{\partial x} =$$

$$\frac{\partial\left(\eta\frac{\partial\psi}{\partial x}\right)}{\partial y} - \frac{\partial\left(\eta\frac{\partial\psi}{\partial y}\right)}{\partial x} = \frac{\partial\left(\psi\frac{\partial\eta}{\partial y}\right)}{\partial x} - \frac{\partial\left(\psi\frac{\partial\eta}{\partial x}\right)}{\partial y} \quad (35)$$

Applying centered second-order finite differencing to these three terms he generated three Jacobian operators and he combined them linearly with arbitrary coefficients. Considering enstrophy conservation by using $\eta J(\psi, \eta)$, he discovered that the sum of all the grid point values over the domain vanished if he weighted the Jacobians equally and gave each a value of one-third. This computational method also assured the conservation of vorticity and energy. The method was ultimately extended to the primitive equations, which represent general three-dimensional flows with divergence. Interestingly, this procedure also prevents nonlinear instability. Yet despite its success, the method implies the solution to an integral by adjustment of the integrand to satisfy a few given constraints, and this may be accomplished in many different ways. Since no proof of optimization has been established, significant truncation errors may occur using this technique.

On the other hand, using the BVE in spectral form, it has been demonstrated both by Platzman (1960) and Lorenz (1960) that the truncated form of the spectral equations as represented by (24) including the truncation of the nonlinear terms as indicated, do maintain these conservation conditions and do not require a special process as is the case for the finite difference equations. This is of course not true for the more general truncated primitive equations, but research with the shallow water equations indicates that the errors are small (Weigle, 1972).

Polar problem

Models represented by finite differences require grids, and the grids for these models have been selected in a variety of ways. The projection equations are gridded to establish some uniformity of scale and allow for the necessary differencing. Since the prediction equations can be represented in spherical coordinates directly on a global surface, gridding the surface in latitude and longitude coordinates seems both more appropriate and simple, and since no mapping is required, this representation has become popular. Unfortunately as one approaches the pole in this representation, the minimum increment of longitude ($\Delta\lambda$) remains the same but the length of the increment decreases with the cosine of latitude. As the linear (CFL) stability criterion depends on the grid length, shorter time steps are required to maintain stability near the pole and leads to the classic 'pole problem'. Various solutions to this problem have been proposed but none are ideal, as the following example suggests. Consider a representation of any model variable where x describes latitude and M refers to half the number of grid points. If one truncates this series so that the terms following N where N>M are removed, the resulting series will represent only larger space scales and a longer time increment will yield stable results. This process has been used frequently. Unfortunately arbitrary truncation of terms in the prediction system may alter the ultimate solution, noting the fragile nature of nonlinear systems.

In the spectral system, scales are directly represented by expansion in wave number, and no special truncation is required to avoid linear stability anywhere on the domain. In addition, all the expansion functions selected are regular at the pole, so no instability can occur due to function convergence problems.

3. SCALING

Scaling is often used as a prelude to truncation. In the physical space domain used by the finite difference system, scales are determined by grid elements with length Δx. In the spectral domain, scales are determined by wave numbers; i.e., they are the transforms of the gridded lengths.

3.1 Scaling in the horizontal domain

Let us first find a characteristic scale in a horizontal two dimensional spectral domain. Considering the properties of the Laplace operator from scale analysis,

$$\nabla^2 \sim \left(\frac{1}{\Delta s}\right)^2 \sim \left(\frac{l}{s}\right)^2 \qquad (36)$$

where l is an index representing the number of sub-regions in s^2, the area of the domain.

Cartesian domain

Assume a function $f(r) = \exp(i\mathbf{k}\cdot\mathbf{r})$ such that,

$$\nabla^2 f = -k^2 f = \left(k_x^2 + k_y^2\right) f \qquad (37)$$

so that $l^2 = k^2 s^2$, a two dimensional index, depends on both horizontal dimensions.

Spherical surface:

Assume a function $f(r) = Y_\alpha(\lambda,\mu) = \exp(im_\alpha\lambda)P_\alpha(\mu)$ such that,

$$\nabla^2 f = \nabla^2 Y_\alpha = n_\alpha(n_\alpha + 1)Y_\alpha \qquad (38)$$

so that $l^2 = n_\alpha(n_\alpha + 1)s^2$. Here the two dimensional index depends only on a single scale index n_α, and linearly if $n_\alpha >> 1$. The scale dependence on the indices in both domains (the Cartesian and the spherical) can be seen from Figure 3. The figure depicts the scaling indices *k* and *n* as they depend on each of the one dimensional indices. The structure of the two dimensional scale elements are obviously rectangles in the Cartesian domain, but are somewhat more complex in the spherical domain. Figure 1 gives a sense of the structure of the scale elements over the sphere for a selected few indices.

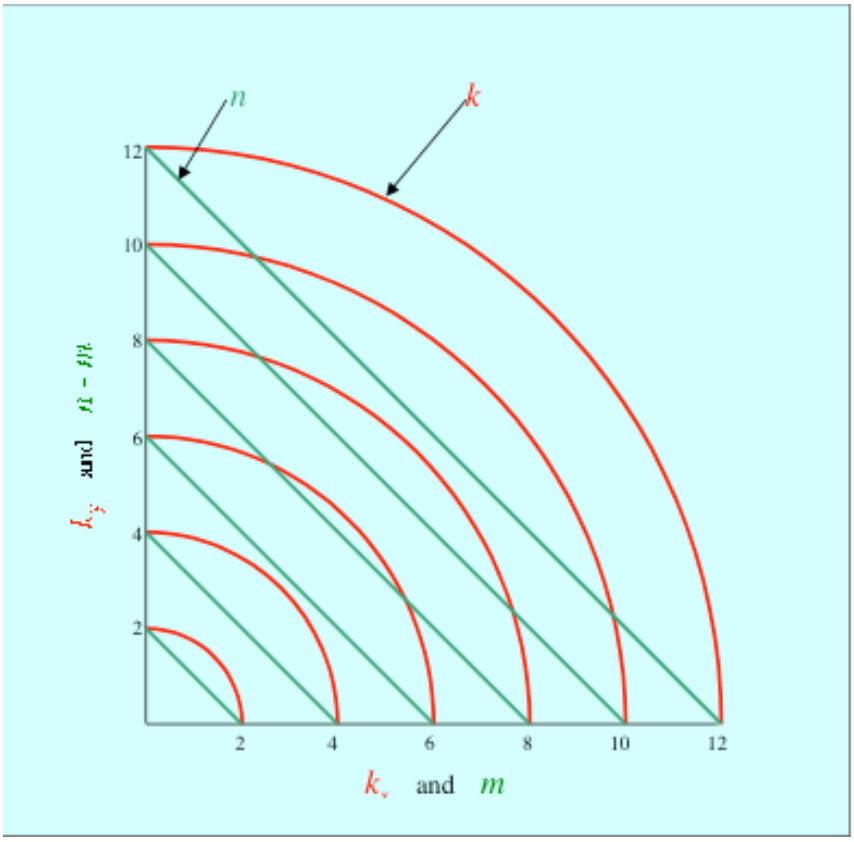

Figure 3. Dependence of the two dimensional indices for the Cartesian and Spectral representations on the one dimensional components of the indices.

3.2 Scaling in the vertical domain

In the finite-difference domain there is no general scaling information. One can get some insight from vertical structure functions taken from data using empirical orthogonal functions (EOFs). Alternately, using a simplified equation for atmospheric flow which has linear solutions, the *quasi-potential vorticity equation*, one has

$$\left(\frac{\partial}{\partial t}+\mathbf{V}\cdot\nabla\right)L(\psi)+\beta\frac{\partial\psi}{\partial x}=0 \qquad (39)$$

and

$$L(\psi)=\nabla^2\psi+\frac{f_o^2\partial}{\overline{\rho}\partial z}\left(\frac{\overline{\rho}\partial\psi}{N^2\partial z}\right) \qquad (40)$$

Vertical structures can be constructed from this vertical equation using the L operator with eigenvalues which are the vertical scaling numbers h_k(*equivalent depths)*,

$$\frac{f_o^2\partial}{\overline{\rho}\partial z}\left(\frac{\overline{\rho}\partial\psi}{N^2\partial z}\right)=\frac{f_o^2}{gh_k} \qquad (41)$$

For simple thermodynamic profiles and BCs, these structures are Bessel functions.

Three-dimensional scaling

In this case we again use the *quasi-potential vorticity equation* as our model. Now the operator L has eigenvalues which depend on the horizontal index *n* and equiv. depth h_k,

$$s^2(n,k)=n(n+1)+\frac{a^2f_o^2}{gh_k} \qquad (42)$$

and these indices can be used for three-dimensional scaling. Table 1 summarizes some values for s as a function of n and h_k.

TABLE 1. Three-dimensional truncation limits on the vertical mode K and on the corresponding horizontal resolution N as determined by the three-dimensional resolution limit S. Here H is the equivalent depth corresponding to the listed vertical mode K in meters, and S is the three-dimensional scaling index of the chosen truncation limit.

K	$H(K)$	N	S	K	$H(K)$	N	S
2	2247.10	3	3.99	26	5.02	84	84.39
3	702.49	6	7.14	27	4.63	87	87.87
4	310.52	10	10.73	28	4.34	90	90.79
5	183.35	13	13.97	29	4.05	93	94.02
6	125.55	16	16.88	30	3.75	97	97.69
7	88.06	20	20.15	31	3.48	101	101.35
8	62.97	23	23.83	32	3.28	104	104.48
9	47.84	27	27.34	33	3.10	107	107.45
10	38.76	30	30.38	34	2.90	110	111.00
11	31.84	33	33.51	35	2.72	114	114.73
12	26.09	36	37.03	36	2.56	118	118.13
13	21.52	40	40.77	37	2.44	121	121.03
14	18.48	43	44.00	38	2.31	124	124.33
15	16.20	46	46.99	39	2.18	128	128.05
16	14.11	50	50.36	40	2.06	131	131.65
17	12.23	54	54.07	41	1.97	134	134.71
18	10.78	57	57.59	42	1.89	137	137.75
19	9.75	60	60.57	43	1.79	141	141.35
20	8.80	63	63.75	44	1.70	145	145.06
21	7.89	67	67.34	45	1.62	148	148.38
22	7.08	71	71.06	46	1.56	151	151.29
23	6.49	74	74.24	47	1.50	154	154.66
24	6.00	77	77.20	48	1.43	158	158.39
25	5.50	80	80.67				

TRUNCATION

Truncation is essential to close a model so that it can be integrated. There are a variety of ways to do this. With the finite difference method, there are many alternatives depending on grid dimensions, point distribution, etc. Things are much more precise for the spectral method.

Two-dimensional truncation with the spectral model

Some options follow:

a) One can truncate at a specified wave number $\alpha = (n + im)_{max}$.

b) One can truncate at a specified ordinal index, n_{max}, and $m \leq n_{max}$; this is denoted as *triangular truncation.*

c) One can truncate at m_{max} and $n \leq n + m_{max}$; this is denoted as *rhomboidal truncation*.

These truncations are depicted graphically on Figure 2 which is a conventional spectral diagram using the indices n and m for coordinates.

Options (b) and (c) have been most popular. It is seen that the energy in atmospheric surfaces converges rapidly with both these truncations. With

reference to relative advantages, (c) has the benefit of giving each planetary wave equal resolution, but (b) is simpler and perhaps more computationally economical, and depends on a single scale index. Additionally, some observational data seems to distribute along lines of constant n. This can be seen on Figure 4, which shows the average KE taken from a data archive on a spectral diagram.

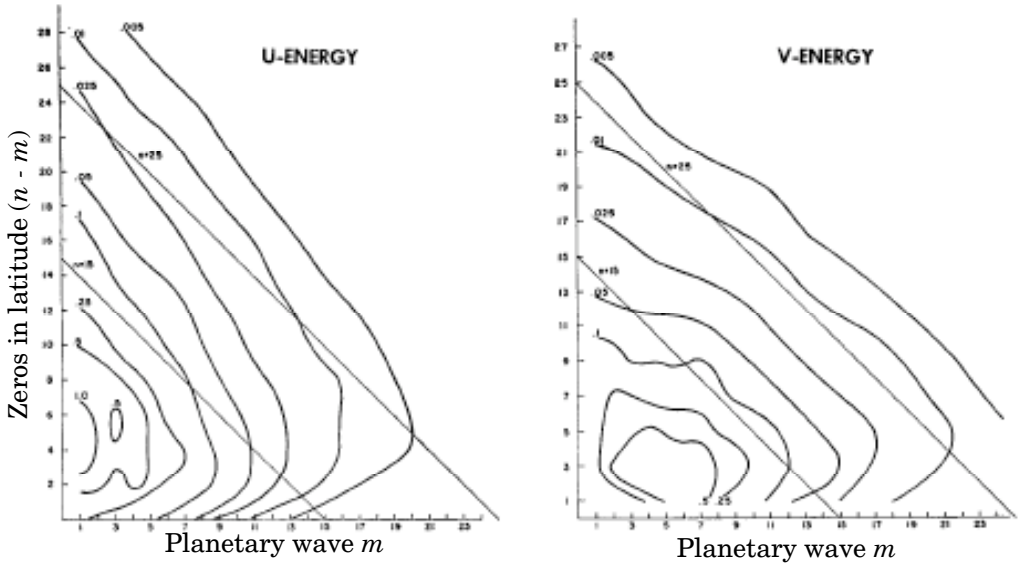

Figure 4. Average KE (both ht. and time) as percent of total in each wave component $\alpha = n + im$ taken from data archives of midlevel atmospheric fields.

Three-dimensional truncation with the spectral model

One can truncate a three-dimensional spectral model using fixed values of the three-dimensional index s (42). In this situation, n is the scaling index in the horizontal domain and k is the index in the vertical domain. Figure 5 gives a graphical representation of some available choices.

As an alternative procedure to selecting three-dimensional truncation for a model, one can find vertical levels in the atmosphere on which the numerical structures of the quasi-geostrophic potential vorticity equation converge to their exact solutions. These levels can be used as optimum vertical levels in a vertical finite difference representation of a spectral model. Thus one can establish a truncated three dimensional model using Legendre polynomials in the horizontal and either vertical structure functions or the corresponding appropriate grid points as described above. Some optimized levels as a function of the total number of levels chosen a model using this technique are described on Table 2. Both of these techniques need careful testing to assess their value in spectral modeling.

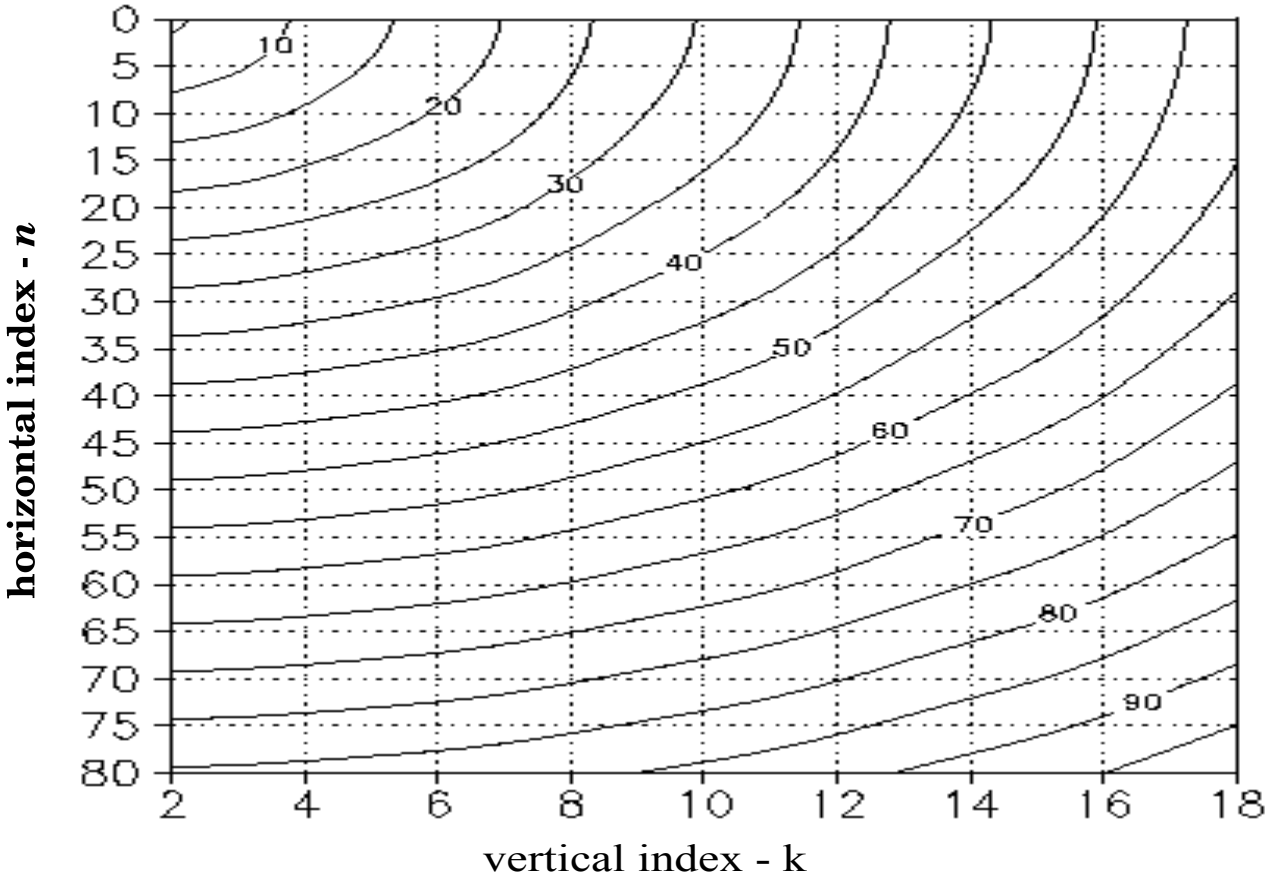

Figure 5. For three-dimensional truncation, lines of constant *s* for various combinations of *k* and *n*.

Number of levels			
7	**9**	**12**	**15**
14.7	13.2	12.3	11.8
		18.1	16.0
	22.3	26.6	21.8
28.3			29.6
	37.1	39.0	40.3
54.6	61.9	57.3	54.8
		84.1	74.4
106.	103.	123.	101.
			138.
	172.	181.	187.
218.	297.	266.	254.
		492.	346.
592.	653.		540.
		819.	827.
943.	963.	985.	986.

Table 2. Optimized σ-levels in hpa as a function of the total number of levels selected.

4. THE COMPUTATIONAL WORLD AND SPECTRAL MODELS

When computational models first appeared, only the computation cycle speed and memory requirements were considered relevant to the integrations. However as computing machines with innovations such as vector processing and later multiprocessors appeared, new issues evolved. Models also became larger such that current models need more computing power than is available. Thus one must use the newer machines most efficiently, and this process takes careful study.

To demonstrate this evolution, consider the following example**.** Define a "computing cycle" to be that time required to do once all calculations that are systematically repeated (in time) to complete the entire calculation. For conventional marching problems this computing cycle is one complete time step. On a serial machine the computing cycle is the total time for that operation and can only be reduced by a faster machine. On a MPP the computing cycle is reduced insofar as many computations can be performed simultaneously. The time required to complete this cycle should converge to the time needed by the machine to perform one computation (the machine cycle) as the number of processors is increased.

We have used this idea on a parallel processor to avail ourselves of the

computing efficiency it afforded. In particular we chose a SIMD parallel processor with 64000 simple processors called the CM-5 and applied the calculation to the spectral (BVE) using the expansion coefficient method. The equations are given by (24) and (25). Note that the tendency of each expansion coefficient ($\psi_\alpha(t)$) is determined by the quadratic product of many such coefficients multiplied by a unique interaction coefficient ($I_{\alpha,\beta,\gamma}$). We distribute the interaction coefficients one to a processor, if there are a sufficient number of them. We then send the appropriate expansion coefficient couples ($\psi_\beta\psi_\gamma$) to the processors, and perform a simple multiplication of all the products in one machine cycle (for as many processors as are available). Note that in the limit, this process can approach one machine cycle. The products are then swept and added as required to form the tendencies of the expansion coefficients, which are then extrapolated in time. This cycle is repeated until the prediction is complete. It should also be noted the interaction coefficient scheme is a highly redundant computation method on a serial processor and is thus computationally inefficient.

We calculated the BVE on a CM-5 using this scheme and compared the results to those retrieved from an identical calculation using the transform method, and discovered that both models ran with comparable speed. Additionally, we used a two-level baroclinic model and ran the same comparison. The results of integrations with those versions of the model also ran with comparable speed. The implication of this experiment suggests that SIMD machines with multi-million processors (a rather inexpensive development) could give the interaction coefficient technique a significant boost in computing efficiency when compared to other methods. This experiment highlights the potential of the spectral method if the appropriate computing machinery is available.

6. HISTORY

Since the 1960s, the spectral method has become by far the most popular technique for converting the prediction equations to a computational form. It appears to overcome many of the limitations introduced by the finite difference method, and despite new ideas that are drawing modelers to other procedures, it remains an attractive method to the modeling community. Perhaps the first successful primitive equation spectral model may be attributed to Bourke [34]. Since that time, most prediction centers have adopted the method. The Canadians and Australians implemented the method in 1976, the National Meteorological Center of NOAA did so in 1980, the French in 1982 and the ECMWF in 1983. As an example of how computing power has evolved, production spectral models at ECMWF have grown in resolution from T63 in 1983 to T213 in 1998 with experiments running at T319.

7. SUMMARY

NWP has made dramatic strides over the last half century following the development of digital computers. No small part of this process is based on the application of computing methods and a thorough knowledge of their features. The utilization of the finite difference method with the prediction equations began the process because the method

had been studied and appeared appropriate to the topic. Concurrently studies began with the spectral method, a procedure not yet at the time in the forefront of competitive methods. The breakthrough came when the transform method was unveiled, because it allowed the number of calculations needed to make a forecast to compete favorably with the finite difference method. As popularity of the spectral method evolved over the years, a number of its features came to light. Those features suggested that the method could produce forecasts more efficiently and perhaps with occasional improvements over other methods.

As noted, the transform method was the single most important feature, which gave the spectral method its impetus. Additionally, it tends to eliminate linear instability linear and phase errors, it bypasses an important nonlinear instability, it conserves important parameters of the system, it avoids the pole problem, it has clear scaling properties which allow for efficient and appropriate truncation based on the important scales that need prediction, and it gives insight on three dimensional truncation wherein the scales in the horizontal and vertical are balanced. From the computational point of view, it allows for efficient computing on parallel processors, thereby speeding the time needed for a forecast. This is becoming increasingly more important as the physics incorporated in models becomes more complex.

Clearly the spectral method has some desirable features that have made it the method of choice over the last decades, but it would be presumptuous to say that NWP would not have advanced to its current position without it. Only time will tell what the optimum method for NWP will be, if there is one.

REFERENCES

Arakawa, A., 1966: Computational design for long-term numerical integrations of the equations of atmospheric motion, *J. Comput. Phys*. **1**, 119-143.

Lorenz, E. N., 1960: Maximum simplification of the dynamic equations, *Tellus* **12**, 243-254.

Phillips, N. A., 1959: An example of non-linear computational instability. The atmosphere and the sea in motion. Rockefeller Institute Press, 501-504.

Platzman, G. W., 1960: The spectral form of the vorticity equation, *J. Meteor* **17**, 635-644.

Richardson, L. F., 1922: *Weather prediction by numerical process*. Cambridge Univ. Press, London and New York.

Weigle, W. F., 1972: *Energy conservation and the truncated spectral form of the primitive equations for a one-layer fluid*. Ph.D. thesis, the University of Michigan, 229pp.